

VU Research Portal

Measuring activity limitations in walking: development of a hierarchical scale for patients with lower-extremity disorders who live at home

Roorda, L.D.; Roebroek, M.E.; van Tilburg, T.G.; Molenaar, I.W.; Lankhorst, G.J.; Bouter, L.M.

published in

Archives of Physical Medicine and Rehabilitation
2005

DOI (link to publisher)

[10.1016/j.apmr.2005.06.014](https://doi.org/10.1016/j.apmr.2005.06.014)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Roorda, L. D., Roebroek, M. E., van Tilburg, T. G., Molenaar, I. W., Lankhorst, G. J., & Bouter, L. M. (2005). Measuring activity limitations in walking: development of a hierarchical scale for patients with lower-extremity disorders who live at home. *Archives of Physical Medicine and Rehabilitation*, 86(12), 2277-2283. <https://doi.org/10.1016/j.apmr.2005.06.014>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

ORIGINAL ARTICLE

Measuring Activity Limitations in Walking: Development of a Hierarchical Scale for Patients With Lower-Extremity Disorders Who Live at Home

Leo D. Roorda, MD, PT, Marij E. Roebroek, PhD, Theo van Tilburg, PhD, Ivo W. Molenaar, PhD, Gustaaf J. Lankhorst, MD, PhD, Lex M. Bouter, PhD, and the Measuring Mobility Study Group

ABSTRACT. Roorda LD, Roebroek ME, van Tilburg T, Molenaar IW, Lankhorst GJ, Bouter LM, and the Measuring Mobility Study Group. Measuring activity limitations in walking: development of a hierarchical scale for patients with lower-extremity disorders who live at home. *Arch Phys Med Rehabil* 2005;86:2277-83.

Objective: To develop a hierarchical scale that measures activity limitations in walking in patients with lower-extremity disorders who live at home.

Design: Cross-sectional study.

Setting: Orthopedic workshops and outpatient clinics of secondary and tertiary care centers.

Participants: Patients (N=981; mean age \pm standard deviation, 58.6 ± 15.4 y; 46% men) living at home, with different lower-extremity disorders: stroke, poliomyelitis, osteoarthritis, amputation, complex regional pain syndrome type I, and diabetic and degenerative foot disorders.

Interventions: Not applicable.

Main Outcome Measures: (1) Fit of the monotone homogeneity model, indicating whether items can be used for measuring patients; (2) fit of the double monotonicity model, indicating invariant (hierarchical) item ordering; (3) intratest reliability, indicating repeatability of the sum score; (4) robustness, addressing the clinimetric properties within subgroups of patients; and (5) differential item functioning, addressing the validity of comparisons between subgroups of patients.

Results: Thirty-five of 41 dichotomous items had (1) good fit of the monotone homogeneity model (coefficient $H=.50$), (2) good fit of the double monotonicity model (coefficient $H^T=.33$), (3) good intratest reliability (coefficient $\rho=.95$), (4) satisfactory robustness (within subgroups of patients defined by age, sex, and diagnosis), and (5) some differential item functioning (6 items in amputees compared with nonamputees).

Conclusions: A hierarchical scale, with excellent scaling characteristics, was developed to measure activity limitations in walking in patients with lower-extremity disorders who live at home. The measurements should be interpreted cautiously when making comparisons between amputees and nonamputees.

Key Words: Activities of daily living; Disability evaluation; Lower extremity; Psychometrics; Questionnaires; Rehabilitation.

© 2005 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

WALKING HAS AN IMPORTANT role in maintaining mobility and independence. However, activity limitations in walking are prevalent, especially among patients with lower-extremity disorders and in the elderly. In The Netherlands' noninstitutionalized population, 10% of the general population and 57% of the population aged 85 years or older have reported limitations in walking.¹ However, little is known about the actual limitations in walking that are perceived by patients who live at home. Furthermore, the severity of these limitations in subgroups of patients, and the determinants of activity limitations in walking, are unclear. A prerequisite in addressing these research questions is a suitable measurement instrument.

Most of the existing instruments do not provide a detailed measurement of limitations in walking as perceived by home-dwelling patients. Many generic²⁻⁷ and disease-specific measurement instruments⁸⁻¹² do provide a measurement of limitations perceived by such patients, but they do not provide a detailed measurement of limitations in walking. Some instruments do provide such a measurement, but those measurements usually result from a patient's timed performance on a test.¹³⁻³¹ Studies of the relation between a patient's test performance and self-reported perception of activity limitations have generally found only low-to-moderate correlations.^{20,32-39} More recently, Stratford et al³⁹ have questioned the content validity of timed performance tests.

Because none of the existing instruments provide a detailed measurement of activity limitations in walking, we developed a new scale that: (1) is suitable for the assessment of patients with lower-extremity disorders who live at home; (2) provides a detailed measurement of activity limitations in walking; (3) is discriminative, which means that it can be used for measuring cross-sectional differences in limitations in walking between patients or subgroups of patients; and (4) is hierarchical, which implies that it consists of items indicating different levels of severity in activity limitations.

This article reports on the development and testing of a detailed and hierarchical scale to measure activity limitations in walking in patients with lower-extremity disorders who live at home.

METHODS

Instrument Development

To create the new instrument, we extracted items we found in an extensive literature review on the assessment of walking. We subjected a first draft version of the instrument to the opinions of

From the Department of Rehabilitation Medicine (Roorda, Lankhorst) and Institute for Research in Extramural Medicine (Roorda, Lankhorst, Bouter), VU University Medical Center, Amsterdam; Department of Rehabilitation Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam (Roebroek); Faculty of Social Sciences, Vrije Universiteit, Amsterdam (Van Tilburg); and Department of Statistics and Measurement, University of Groningen, Groningen (Molenaar), The Netherlands.

Supported by the SGO Health Research Promotion Program.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit on the author(s) or on any organization with which the author(s) is/are associated.

Reprint requests to Leo D. Roorda, MD, PT, VU University Medical Center, Dept of Rehabilitation Medicine, PO Box 7057, 1007 MB Amsterdam, The Netherlands, e-mail: ld.roorda@vumc.nl.

0003-9993/05/8612-9853\$30.00/0

doi:10.1016/j.apmr.2005.06.014

experts (physicians, physical and occupational therapists, sociologists), and pretested this version. We tested a second version in 345 patients with orthopedic and rheumatologic disorders of the lower extremity,⁴⁰ which resulted in the rewording of some items. The final version of the instrument that we tested in the study we report here was a self-administered questionnaire with 41 items. Appendix 1 provides a summary of the instructions for the patients and lists the 41 items. The items operationalize 9 aspects of walking: (1) distance (items 1–4, 20), (2) time (items 5, 8, 21, 24), (3) velocity (items 6, 13, 22, 29, 37, 38), (4) frequency (items 7, 14, 23, 28, 30), (5) adaptations (items 9, 12, 15, 25, 31), (6) difficulty (items 10, 16, 26, 32, 39), (7) uncertainty (items 11, 17, 27, 33), (8) use of aids (items 18, 34, 35, 40), and (9) use of help (items 19, 36, 41). The items are written in behavioral terms and refer to what patients actually do, not to what they think that they can do. We chose dichotomous response options (“yes” box marked, “yes” box not marked) to facilitate interpretation.

Participants

To test the instrument, we recruited patients from several studies of subjects with lower-extremity disorders who lived at home. First, we sampled patients in a community-based cohort study who had chronic stroke,⁴¹ at the 2-year follow-up measurement. Second, we sampled poliomyelitis patients from a cohort study,⁴² also at the 2-year follow-up measurement. Third, we sampled outpatients with hip or knee osteoarthritis (OA) who were undergoing arthroplasty at 2 general hospitals and 1 university hospital in a ongoing inception cohort study, directly after placement on the waiting list.⁴³ Fourth, we sampled outpatients with lower-extremity amputation who were in an ongoing inception cohort study, at the completion of their multidisciplinary rehabilitation treatment. In addition, we sampled consecutive eligible outpatient amputees from 2 orthopedic workshops. Fifth, we sampled patients with lower-extremity complex regional pain syndrome type I (CRPS I), who were participating in a randomized controlled trial being conducted by the anesthesiology and surgical outpatient clinics of 2 university hospitals,⁴⁴ at the baseline measurement. We also sampled consecutive eligible patients with CRPS I from the same departments.⁴⁵ A sixth sample was of consecutive eligible patients with diabetic foot disorders (with and without foot ulcers) who were in an ongoing case-control study in a rehabilitation center’s outpatient department.⁴⁶ We also sampled consecutive eligible outpatients with diabetic foot disorders who were receiving podiatric care at a general and at a university hospital in an observational study,⁴⁷ at the baseline measurement. Finally, we sampled patients with degenerative foot disorders who were supplied with foot orthoses from orthopedic workshops in an observational study.⁴⁸ All patients completed the questionnaire and provided additional information about their age and sex.

Analysis

Scalability and invariant item ordering. Scalability of an item set implies that it can be used for the measurement of patients, whereas invariant item ordering implies that the (hierarchical) ordering of the items is the same for all patients. We used Mokken scale analysis to investigate the scalability and the ordering of the items in this study.^{49–52} Mokken scale analysis is a nonparametric approach to item response theory (IRT). According to the IRT, the abilities of patients are latent traits. Mokken scale analysis provides only ordinal information about the location of patients and items on the scale of the latent trait. Patients are ordered on this scale according to their sum scores. Patients with higher sum scores have more activity

limitations. Items are ordered on the scale of the latent trait according to their mean score, which is the proportion of patients responding positively to the item at issue. Items with higher mean scores indicate less activity limitations. Within the framework of Mokken scale analysis, the scalability of an item set is studied by investigating the fit of the monotone homogeneity (MH) model, whereas the invariant ordering of the items is studied by investigating the fit of the double monotonicity (DM) model.

Evaluation of the MH model. The fit of the MH model is evaluated by calculating the scalability coefficient H ,^{50,51} which is a global indicator of the degree to which patients can be accurately ordered on the latent trait by means of their sum score. Scale criteria are met when (1) the coefficients of scalability for all item pairs (H_{ij}) are positive, (2) the scalability coefficients for the items in relation to the scale at issue (H_i) have values of least .30, and (3) the scalability coefficient for the scale (H) is at least .30. Higher values for H_i and H imply a better scale. Generally, a scale is considered to be strong when H values are equal to or exceed .50, moderate when values are from .40 to .50, and weak when values are from .30 to .40.^{50,51}

Evaluation of the DM model. Evaluation of the fit of the DM model with the data starts by investigating the fit of the MH model and continues by calculating the coefficient H^T ,^{50,51} which is a global indicator of the degree to which the ordering of the items is invariant across the latent trait. Criteria for invariant item ordering are met when the percentage of negative coefficients at the level of the individual patients (H^T_i) is less than 10, and the coefficient for the total set of patients (H^T) is at least .30.^{50,51} The larger the H^T , the greater the confidence that can be assigned to the invariant ordering of items across the latent trait.

Intratest reliability. The intratest reliability (or internal consistency) concerns the degree of repeatability of the sum score. We quantified the intratest reliability by calculating the reliability coefficient ρ .^{50,51} A reliability coefficient of .90 or more is recommended for decisions about individual patients.⁵³

Robustness. Robustness concerns the scalability of the item set and the invariant item ordering within subgroups of patients. In this study we investigated robustness with respect to age, sex, and diagnosis. With respect to age, we dichotomized the study population on the basis of its median age of 59.9 years. Because good estimates of scalability require large groups of patients (not less than 200 and preferably at least 500),⁵⁰ we investigated robustness with respect to diagnosis only for the subgroups of patients with OA, amputation, and degenerative foot disorders, and for a rest group consisting of patients with the other diagnoses. We evaluated scalability and invariant item ordering within subgroups of patients, as described, for all patients.

Differential item functioning. We also investigated differential item functioning (DIF), or item bias, for age, sex, and diagnosis in the same patient subgroups as above. DIF addresses the issue of making valid comparisons between subgroups of patients, for example, between OA patients and amputees. Suppose that a certain item functions differently in those patients, for instance, that it indicates less activity limitations in amputees than in OA patients, but otherwise patients in both subgroups have the same amount of limitations, according to the other items. In such a case, amputees will more often respond positively to this item, and they will tend to have higher scores than the OA patients. So, an item that functions differently in subgroups of patients causes differences in subgroup scores, even when the subgroups have a similar severity of activity limitations. As a consequence, DIF impedes valid comparisons between these subgroups. DIF is visualized in a scatter plot.

RESULTS

Participants

We enrolled 1043 participants in this study: 781 with response rates (participants/participants eligible at the time of measurement) and 262 without. The 262 participants without response rates consisted of consecutive eligible amputees ($n=169$), patients with CRPS I ($n=29$),⁴⁵ and patients with diabetic foot disorders ($n=64$).^{46,47} The response rate of the other 781 participants was 85% (781/922), and for the different diagnostic groups it was as follows: stroke, 96% (73/76)⁴¹; poliomyelitis, 87% (90/103)⁴²; hip OA, 89% (211/238)⁴³; knee OA, 91% (84/92); transfemoral or knee disarticulation amputation, 100% (34/34); transtibial amputation, 98% (53/54); other amputation, 100% (10/10); CRPS I, 100% (41/41)⁴⁴; and degenerative foot disorders, 68% (185/274).⁴⁸

Of the 1043 patients enrolled in our study, we excluded from the scale analysis 62 who stated that they did not walk inside or outside the home at all. As a consequence, we included data on 981 patients in the scale analysis. Their mean age \pm standard deviation was 58.6 ± 15.4 years, and 46% ($n=453$) were men. They had the following diagnoses: stroke ($n=67$); poliomyelitis ($n=89$); hip OA ($n=201$); knee OA ($n=80$); transfemoral or knee disarticulation amputation ($n=105$); transtibial amputation ($n=114$); other amputation ($n=20$); CRPS I ($n=65$); diabetic foot disorders ($n=57$); and degenerative foot disorders ($n=183$).

Mokken Scale Analysis

MH model. The scale criteria were met for 35 of the 41 items (table 1), which indicates that those 35 items can be used for the measurement of patients. Their coefficients of scalability for the item pairs (H_{ij}) were positive, the scalability coefficients for the items in relation to the scale at issue (H_i) were at least .35, and the scalability coefficient of the scale (H) was .50, which indicates a strong scale. The reasons for removing from the scale 6 of 41 items were negative H_{ij} (see appendix 1, item 19) or dubious fit according to H_i (see appendix 1, items 3, 18, 28, 36, 41).

DM model. The criteria for invariant item ordering were also met by the 35 items, which indicates that the hierarchical ordering, from indicating minor limitations to indicating severe limitations, was the same for the entire range of patients. The percentage of negative coefficients at the level of the individual patients (H_a^T) was 4%, while the coefficient for the total set of patients (H^T) was .32 (see table 1).

Responding positively to an item about activity limitations in walking inside the house indicated more limitations than responding positively to an item about such limitations outside the house. The 35-item set contains 13 pairs of items that differ only with respect to walking inside or outside the house, for example, such as items 4 and 20. For 12 of these 13 pairs of items, the item about walking inside the house had a lower mean score, indicating more limitations than the corresponding item about walking outside the house.

Intratest reliability. The intratest reliability coefficient ρ was .95 for the 35 items, which indicates that their intratest reliability was good enough for making decisions about individual patients.

Robustness. Robustness with respect to age, sex, and diagnosis was satisfactory (table 2), which indicates that this 35-item set can be used for measurement within the subgroups of patients, and that the item ordering is invariant within these subgroups. In the subgroup of OA patients, we found a scalability coefficient H of .37, indicating a weak scale, and 6 items

Table 1: Coefficient of Scalability of the 35 Items, and Coefficient of Scalability, Invariant Item Ordering, Intratest Reliability, and Median (IQR) Sum Score of the Scale of Activity Limitations in Walking for All Patients

Item*	Mean [†]	H_i^{\ddagger}
Minor limitations [§]		
More slowly (37)	.70	.71
Outside: shorter distances (20)	.66	.67
Outside: more slowly (22)	.58	.64
Outside & obstacles: more slowly (29)	.57	.66
Outside: shorter periods (21)	.45	.56
Outside: with difficulty (26)	.45	.56
Inside: different way (9)	.43	.50
Outside: different way (25)	.42	.48
Crossroads: takes longer (38)	.41	.51
Inside: more slowly (6)	.38	.51
Outside: less often (23)	.38	.50
Outside & obstacles: with difficulty (32)	.37	.55
Inside & obstacles: more slowly (13)	.36	.51
Outside: with an aid (34)	.36	.40
Inside: with difficulty (10)	.35	.53
Outside & obstacles: less often (30)	.34	.51
With an aid (40)	.33	.38
Outside: stand still more often (24)	.32	.45
Outside: unsteadily (27)	.31	.47
Outside & obstacles: unsteadily (33)	.30	.49
Outside: distances with an aid (35)	.30	.38
Outside & obstacles: different way (31)	.26	.49
Inside & obstacles: with difficulty (16)	.25	.50
Crossroads: with difficulty (39)	.24	.48
Inside: unsteadily (11)	.23	.48
Inside: shorter periods (5)	.20	.47
Inside: less often (7)	.20	.48
Inside & obstacles: unsteadily (17)	.20	.49
Inside & obstacles: different way (15)	.18	.48
Inside: stand still more often (8)	.17	.48
Inside: hold on to something (12)	.16	.43
Inside: shorter distances (4)	.14	.43
Inside & obstacles: less often (14)	.14	.51
Inside: don't come into all rooms (2)	.08	.35
Inside: in one room (1)	.07	.38
Severe limitations		
Coefficient of scalability H		.50
Coefficient of invariant item ordering H^T [¶]		.32
Percentage of negative H_a^T values [#]		4
Coefficient of reliability ρ^{**}		.95
Sum score median ^{††}	10	
Sum score IQR	3–18	

Abbreviation: IQR, interquartile range.

*Abbreviated item. Numbers in parentheses refer to the nonabbreviated items in the appendix.

[†]Item mean score, indicating the proportion of patients responding positively to the items. Items with higher mean scores indicate less activity limitations.

[‡]Coefficient of scalability H_i of the items in relation to the scale (range, 0–1 under the MH model). A minimum value of $H_i=.30$ is recommended.

[§]Severity of activity limitations increases in the scale from top to bottom.

^{||}Coefficient of scalability H for the scale (range, 0–1 under the MH model). A scale is considered to be strong when $H \geq .50$.

[¶]Coefficient of invariant item ordering H^T for the total set of patients. A minimum value of $H^T=.30$ is recommended.

[#]Percentage of negative coefficients of invariant item ordering H_a^T at the level of the individual patients. A percentage of <10 is recommended.

^{**}Reliability coefficient ρ for the scale (range, 0–1 under the MH model). A coefficient $\rho \geq .90$ is recommended for decisions about individual patients.

^{††}Sum scores range from 0 to 35. Patients with higher sum scores have more limitations.

Table 2: Coefficients of Scalability, Invariant Item Ordering, Intratest Reliability, and Median (IQR) Sum Scores of the Scale of Activity Limitations in Walking for Subgroups of Patients, Defined by Age, Sex, and Diagnosis

Scale Statistic	Age		Sex		Diagnosis			
	Young (n=491)	Old (n=490)	Men (n=453)	Women (n=528)	OA (n=281)	Amp (n=239)	Foot (n=183)	Other (n=278)
Coefficient of scalability H^*	.54	.45	.52	.48	.37	.49	.43	.50
Coefficient of invariant item ordering H^T	.36	.31	.34	.32	.34	.34	.36	.36
Percentage of negative H^T values [†]	4	4	3	5	5	3	3	3
Coefficient of reliability $\rho^§$.96	.95	.96	.95	.92	.96	.94	.95
Sum score median	7	11	10	9	14	10	2	10
Sum score IQR	1–17	5–20	3.5–19	3–17	9–21	4–20	0–5	3–19.25

Abbreviation: Amp, amputation.

*Coefficient of scalability H for the scale (range, 0–1 under the MH model). A scale is considered to be strong when $H \geq .50$, medium when $.50 > H \geq .40$, and weak when $.40 > H \geq .30$.

†Coefficient of invariant item ordering H^T for the total set of patients. A minimum value of $H^T = .30$ is recommended.

§Percentage of negative coefficients of invariant item ordering H^T at the level of the individual patients. A percentage of <10 is recommended.

||Reliability coefficient ρ for the scale (range, 0–1 under the MH model). A coefficient $\rho \geq .90$ is recommended for decisions about individual patients.

||Sum scores range from 0 to 35. Patients with higher sum scores have more limitations.

(item 1, $H_i = .25$; item 2, $H_i = .27$; item 12, $H_i = .25$; item 34, $H_i = .21$; item 35, $H_i = .27$; item 40, $H_i = .17$) showed some misfit of the MH model ($H_i < .30$). For the subgroup of patients with degenerative foot disorders, 4 items (item 2, $H_i = .26$; item 16, $H_i = .26$; item 34, $H_i = .26$; item 35, $H_i = .25$) showed some misfit of the MH model. The H_i coefficients of the other items within the defined subgroups of patients are available from the authors on request.

Differential item functioning. We found minimal DIF with respect to age, no DIF with respect to sex, and some DIF with respect to diagnosis for the 35 items. The DIF with respect to age involved only item 34, indicating that valid comparisons can be made between younger and older patients. The fact that we found no DIF with respect to sex indicates that valid comparisons can be made between men and women. The DIF with respect to diagnosis involved items 13, 15, 34, 35, and 40, indicating that comparisons between diagnostic groups should be made cautiously. The DIF tended to be mainly attributable to the amputees. Consequently, in an additional analysis, we dichotomized the study population (amputees vs nonamputees) and found DIF with respect to 6 items (items 13, 15, 31, 34, 35, 40) (fig 1). Notice that for amputees, these 6 items all have higher mean scores, which will result in their having higher sum scores. Other ways of dichotomizing the study population with respect to diagnosis—OA versus non-OA and degenerative foot disorders versus nondegenerative foot disorders—revealed no DIF (data not shown).

DISCUSSION

Our objective was to develop and test a measurement instrument to provide a detailed assessment of activity limitations in walking for patients with lower-extremity disorders who live at home. To test the instrument, we sampled only outpatients who had different disorders of the lower extremity, originating from different parts of the body: the brain (stroke), the peripheral nerve (poliomyelitis, diabetic foot disorders), the hip (OA), the upper leg (transfemoral amputation), the knee (OA, knee disarticulation amputation), the lower leg (transtibial amputation), and the foot (foot amputations, CRPS I, diabetic and degenerative foot disorders).

We sampled study populations from several ongoing studies with different study designs and different sampling techniques. This resulted in a difference in the quality of the patient samples we included and raises some concerns with regard to

the groups of consecutive eligible patients—mainly amputees—for whom the selection process is unclear. This may have resulted in selection bias.

We confirmed the suitability of 35 of 41 items for the measurement of patients by demonstrating good fit with the MH model. We excluded 6 items from the scale, mainly because of dubious fit. Three of the 6 concerned the use of help (see appendix 1; items 19, 36, 41). Responding positively to an item about being helped by someone will only partially depend on a patient's activity limitations. Other factors—including the

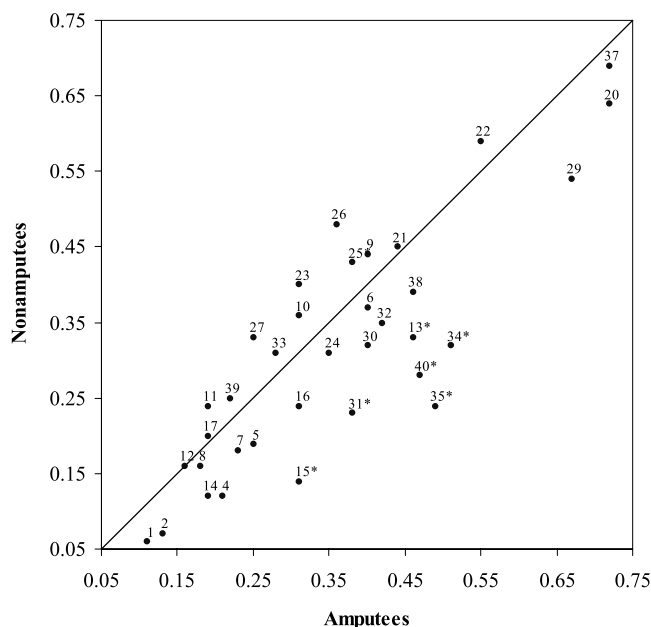


Fig 1. Mean item scores for amputees and nonamputees. Mean item scores for amputees are on the x axis, and for nonamputees on the y axis. Low mean item scores designate the items indicating severe limitations, and high mean item scores the items indicating minor limitations. The numbers near the data points refer to the item numbers. For a full explanation of the items, see appendix 1. An identity line is drawn through the origin with a slope of 1. Items with major and significant deviations from the identity line demonstrate DIF. *Items with DIF.

availability of somebody who can offer this help—will also contribute to the probability of a positive response. If that probability is not just the result of the activity limitations of the patients, an item can demonstrate misfit.

We confirmed invariant item ordering, and thus a hierarchy, by demonstrating good fit of the DM model for all patients and within subgroups of patients. The items can therefore be ordered from those indicating minor limitations to those indicating severe limitations. Because of the invariant item ordering, a patient with a sum score of 4 will generally respond positively to the first 4 items indicating minor limitations (see table 1; items 37, 20, 22, 29) and negatively to the other 31 items that indicate more severe limitations. The finding that items about walking inside the house indicate more activity limitations than items about walking outside the house confirms the results of the study by Van Buuren and Hopman-Rock,⁵⁴ and is in agreement with clinical observations.

We confirmed the suitability of the 35 items for measurement within subgroups of patients—defined by age, sex, and diagnosis—by demonstrating good fit with the MH model within these subgroups. For the subgroups of OA and degenerative foot patients only, 6 and 4 items, respectively, showed misfit of the model. This might be explained for some items (see table 1; OA: items 1, 2; degenerative foot: items 2, 16, 34, 35) by the small number of patients in these subgroups who responded positively to those items. In such cases, violation of the model caused by unusual answers from a few patients can easily result in a low estimate of the H_i coefficient.

We have some concerns about the validity of comparisons between amputees and nonamputees, because 6 of 35 items demonstrate DIF. The finding that amputees more often respond positively to items about walking in a different way over “obstacles” (see appendix 1; items 15, 31) may reflect the fact that unilateral amputees tend to put their nonamputated leg on or over an obstacle first. The finding that amputees respond more often to items about the use of aids (see appendix 1; items 34, 35, 40) may reflect the fact that they use more aids than nonamputees, or that some refer to their prosthesis as an aid. Although items with DIF do not always produce poor measurements,⁵⁵ the use of the 29 items without DIF only can be considered when comparisons are made between amputees and nonamputees.

Future research should focus on the 35 items’ other clinimetric properties that we did not address. For instance, the construct validity of the instrument should be tested. Construct validity indicates that the new instrument relates to other measures, as hypothesized.⁵⁶ Future research should also focus on longitudinal (or evaluative) clinimetric properties, the ceiling and floor effects, and the interpretability of the scores.

In summary, testing the items yielded satisfactory results for 35 of 41 items, and the resulting scale can be recommended as a measurement instrument for use in clinical research. For some applications in daily clinical practice, however, the complete instrument may be too detailed. In such cases only part of the instrument could be used. The items indicating minor activity limitations (see table 1, items 37, 20, 22, 29) could be used for a quick screening. It is unlikely that patients who respond negatively to those items will have activity limitations in walking. We think that this agrees with what is found in clinical practice in rehabilitation medicine. For instance, a physician asks a patient whether he/she has limitations in walking. If the patient responds positively, the physician addresses this topic in more detail and tries to assess the severity of the activity limitation. If the patient responds negatively, the physician will address—after this quick screening—other pos-

sible activity limitations, for instance, limitations in climbing stairs.

CONCLUSIONS

A hierarchical scale with excellent scaling characteristics has been developed to provide a detailed measurement of activity limitations in walking in patients with lower-extremity disorders who live at home. However, the measurements should be interpreted cautiously when comparisons are made between amputees and nonamputees.

Acknowledgments: Members of the Measuring Mobility Study Group who are co-authors of this article are: Anne M. Boonstra, MD, PhD, Rehabilitation Center Revalidatie Friesland, Beetsterzwaag; Fred A. de Laat, MD, Roessingh Research and Development, Enschede, and Rehabilitation Center Tolbrug, ‘s Hertogenbosch; Jacob J. Caron, MD, Department of Orthopedic Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam; Bart J. Burger, MD, PhD, Department of Orthopedic Surgery, Medisch Centrum Alkmaar, Alkmaar; Ide C. Heyligers, MD, PhD, Department of Orthopedic Surgery, VU University Medical Center, Amsterdam; Frans Nollet, MD, PhD, Department of Rehabilitation, Academic Medical Center, University of Amsterdam, Amsterdam; Irene E. Stover-Van Herk, MSc, Northern Center for Health Care Research, University of Groningen, Groningen; Roberto S. Perez, PhD, Department of Anesthesiology, VU University Medical Center, Amsterdam; Jan W. Meijer, MD, PhD, Rehabilitation Center Tolbrug, ‘s Hertogenbosch, and Center for Rehabilitation at Beatrixoord, University Hospital Groningen, Groningen; and P. Mieke Rijken, PhD, Netherlands Institute of Health Services Research (NIVEL), Utrecht, The Netherlands.

APPENDIX 1: SUMMARY OF THE INSTRUCTIONS FOR THE PATIENTS AND THE 41 TESTED ITEMS

Please answer YES to every statement that **both** applies to your current situation and is connected with your health.

1. I walk in one room and not in other rooms (eg, I only walk in the living room or in the bedroom).
2. I walk in the house but I don’t come into all the rooms.
3. I only walk in the house.
4. I walk in the house, but shorter distances.
5. I walk in the house, but for shorter periods.
6. I walk in the house, but more slowly.
7. I walk in the house, but less often.
8. I walk in the house, but I stand still for a moment more often.
9. I walk in the house, but in a different way (eg, I limp, I wiggle, I stumble or I have a stiff leg).
10. I walk in the house, but with (some) difficulty.
11. I walk in the house, but I walk unsteadily.
12. I walk in the house and (almost) always hold on to something (eg, the table, a piece of furniture, or the wall).
13. I walk in the house, but I walk more slowly over “obstacles” (eg, thresholds or steps).
14. I walk in the house, but I less often walk over “obstacles.”
15. I walk in the house, but I walk in a different way over “obstacles” (eg, I pull up one leg at a time).
16. I walk in the house, but with (some) difficulty over “obstacles.”
17. I walk in the house, but I walk unsteadily over “obstacles.”
18. I walk in the house (almost) always with an aid (eg, with a stick, a crutch, a roller, or a frame).
19. I walk in the house and am (almost) always helped by someone.
20. I do walk outside, but shorter distances.

APPENDIX 1: SUMMARY OF THE INSTRUCTIONS FOR THE PATIENTS AND THE 41 TESTED ITEMS (Cont'd)

21. I do walk outside, but for shorter periods.
22. I do walk outside, but more slowly.
23. I do walk outside, but less often.
24. I do walk outside, but I stand still for a moment more often.
25. I do walk outside, but in a different way (eg, I limp, I wiggle, I stumble, or I have a stiff leg).
26. I do walk outside, but with (some) difficulty.
27. I do walk outside, but I walk unsteadily.
28. I do not walk outside at all if the weather is bad (eg, if there is a strong wind or if it is raining).
29. I do walk outside, but I walk more slowly over "obstacles" (eg, steps, curbs, bad roads, or uneven surfaces).
30. I do walk outside, but I walk less often over "obstacles."
31. I do walk outside, but I walk in a different way over "obstacles" (eg, I pull up one leg at a time).
32. I do walk outside, but with (some) difficulty over "obstacles."
33. I do walk outside, but I walk unsteadily over "obstacles."
34. I walk outside (almost) always with an aid (eg, with a stick, a crutch, a roller or a frame).
35. I (almost) always walk longer distances outside with an aid.
36. I walk outside and am (almost) always helped by someone.
37. I walk more slowly.
38. I cross roads, but it takes me longer.
39. I cross roads, but with (some) difficulty.
40. I (almost) always use an aid to keep up with other people (eg, a stick, a crutch, a roller or a frame).
41. I am (almost) always helped by someone to keep up with other people.

References

1. Netherlands Central Bureau of Statistics. Physical disability in the population of the Netherlands 1986/1988. The Hague: SDU Publishers/CBS Publications; 1990.
2. Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. *Md State Med J* 1965;14:61-5.
3. Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health* 1980;34:281-6.
4. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981;19:787-805.
5. De Bruin AF, Diederiks JP, De Witte LP, Stevens FC, Philipsen H. The development of a short generic version of the Sickness Impact Profile. *J Clin Epidemiol* 1994;47:407-18.
6. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil* 1987;1:6-18.
7. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
8. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am* 1969;51:737-55.
9. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *J Rheumatol* 1982;9:789-93.
10. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to anti-rheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833-40.
11. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. *Clin Orthop* 1989;Nov(248):13-4.
12. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. *Arthritis Rheum* 1992;35:1-10.
13. Marks R. Reliability and validity of self-paced walking time measures for knee osteoarthritis. *Arthritis Care Res* 1994;7:50-3.
14. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med* 1995;332:556-61.
15. Ostchega Y, Harris TB, Hirsch R, Parsons VL, Kington R, Katzoff M. Reliability and prevalence of physical performance examination assessing mobility and balance in older persons in the US: data from the Third National Health and Nutrition Examination Survey. *J Am Geriatr Soc* 2000;48:1136-41.
16. Schwartz AV, Villa ML, Prill M, et al. Falls in older Mexican-American women. *J Am Geriatr Soc* 1999;47:1371-8.
17. Ward MM. Clinical measures in rheumatoid arthritis: which are most useful in assessing patients? *J Rheumatol* 1994;21:17-27.
18. Spiegel JS, Paulus HE, Ward NB, Spiegel TM, Leake B, Kane RL. What are we measuring? An examination of walk time and grip strength. *J Rheumatol* 1987;14:80-6.
19. Grace EM, Gerez EM, Kassam YB, Buchanan HM, Buchanan WW, Tugwell PS. 50-foot walking time: a critical assessment of an outcome measure in clinical therapeutic trials of antirheumatic drugs. *Br J Rheumatol* 1988;27:372-4.
20. Kennedy D, Stratford PW, Pagura SM, Walsh M, Woodhouse LJ. Comparison of gender and group differences in self-report and physical performance measures in total hip and knee arthroplasty candidates. *J Arthroplasty* 2002;17:70-7.
21. Reuben DB, Siu AL. An objective measure of physical function of elderly outpatients. The Physical Performance Test. *J Am Geriatr Soc* 1990;38:1105-12.
22. Brown M, Sinacore DR, Binder EF, Kohrt WM. Physical and performance measures for the identification of mild to moderate frailty. *J Gerontol A Biol Sci Med Sci* 2000;55:M350-5.
23. King MB, Judge JO, Whipple R, Wolfson L. Reliability and responsiveness of two physical performance measures examined in the context of a functional training intervention. *Phys Ther* 2000;80:8-16.
24. Mathias S, Nayak US, Isaacs B. Balance in elderly patients: the "get-up and go" test. *Arch Phys Med Rehabil* 1986;67:387-9.
25. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991;39:142-8.
26. Wall JC, Bell C, Campbell S, Davis J. The Timed Get-up-and-Go test revisited: measurement of the component tasks. *J Rehabil Res Dev* 2000;37:109-13.
27. Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. *J Am Geriatr Soc* 1986;34:119-26.
28. Galindo-Ciocon DJ, Ciocon JO, Galindo DJ. Gait training and falls in the elderly. *J Gerontol Nurs* 1995;21:10-7.
29. Tinetti ME, Ginter SF. Identifying mobility dysfunctions in elderly patients. Standard neuromuscular examination or direct assessment? *JAMA* 1988;259:1190-3.
30. Robbins AS, Rubenstein LZ, Josephson KR, Schulman BL, Osterweil D, Fine G. Predictors of falls among elderly people. Results of two population-based studies. *Arch Intern Med* 1989;149:1628-33.

31. Harada N, Chiu V, Damron-Rodriguez J, Fowler E, Siu A, Reuben DB. Screening for balance and mobility impairment in elderly individuals living in residential care facilities. *Phys Ther* 1995; 75:462-9.
32. Elam JT, Graney MJ, Beaver T, el Derwi D, Applegate WB, Miller ST. Comparison of subjective ratings of function with observed functional ability of frail older persons. *Am J Public Health* 1991;81:1127-30.
33. Reuben DB, Valle LA, Hays RD, Siu AL. Measuring physical function in community-dwelling older persons: a comparison of self-administered, interviewer-administered, and performance-based measures. *J Am Geriatr Soc* 1995;43:17-23.
34. Cress ME, Schechtman KB, Mulrow CD, Fiatarone MA, Gerety MB, Buchner DM. Relationship between physical performance and self-perceived physical function. *J Am Geriatr Soc* 1995;43: 93-101.
35. Hoeymans N, Feskens EJ, Van den Bos GA, Kromhout D. Measuring functional status: cross-sectional and longitudinal associations between performance and self-report (Zutphen Elderly Study 1990-1993). *J Clin Epidemiol* 1996;49:1103-10.
36. Finch E, Walsh M, Thomas SG, Woodhouse LJ. Functional ability perceived by individuals following total knee arthroplasty compared to age-matched individuals without knee disability. *J Orthop Sports Phys Ther* 1998;27:255-63.
37. Steultjens MP, Dekker J, Van Baar ME, Oostendorp RA, Bijlsma JW. Internal consistency and validity of an observational method for assessing disability in mobility in patients with osteoarthritis. *Arthritis Care Res* 1999;12:19-25.
38. Steultjens MP, Roorda LD, Dekker J, Bijlsma JW. Responsiveness of observational and self-report methods for assessing disability in mobility in patients with osteoarthritis. *Arthritis Rheum* 2001;45:56-61.
39. Stratford PW, Kennedy D, Pagura SM, Gollish JD. The relationship between self-report and performance-related measures: questioning the content validity of timed tests. *Arthritis Rheum* 2003; 49:535-40.
40. Roorda LD, Roebroeck ME, Lankhorst GJ, Van Tilburg T, Bouter LM. Measuring functional limitations in rising and sitting down: development of a questionnaire. *Arch Phys Med Rehabil* 1996; 77:663-9.
41. Van Herk IE, Arendzen JH, Rispens P. Ten-metre walk, with or without a turn? *Clin Rehabil* 1998;12:30-5.
42. Nollet F, Beelen A, Twisk JW, Lankhorst GJ, De Visser M. Perceived health and physical functioning in postpoliomyelitis syndrome: a 6-year prospective follow-up study. *Arch Phys Med Rehabil* 2003;84:1048-56.
43. Roorda LD, Jones CA, Waltz M, et al. Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. *Ann Rheum Dis* 2004;63:36-42.
44. Perez RS, Zuurmond WW, Bezemer PD, et al. The treatment of complex regional pain syndrome type I with free radical scavengers: a randomized controlled study. *Pain* 2003;102:297-307.
45. Perez RS, Roorda LD, Zuurmond WW, Bannink II, Vranken JH, De Lange JJ. Measuring perceived activity limitations in lower extremity Complex Regional Pain Syndrome type 1 (CRPS I): test-retest reliability of two questionnaires. *Clin Rehabil* 2002;16: 454-60.
46. Meijer JW, Trip J, Jaegers SM, et al. Quality of life in patients with diabetic foot ulcers. *Disabil Rehabil* 2001;23:336-40.
47. Rijken PM, Dekker J, Lankhorst GJ, et al. Podiatric care for diabetic patients with foot problems: an observational study. *Int J Rehabil Res* 1999;22:181-8.
48. Kruizinga CP, Boonstra A, Groothoff JW, Elzinga A, Goeken LN. Characteristics of patients supplied with foot orthoses from orthopaedic workshops in The Netherlands. *Foot* 2003;13:39-45.
49. Mokken RJ. Nonparametric models for dichotomous responses. In: Van Der Linden WJ, Hambleton RK, editors. *Handbook of modern item response theory*. New York: Springer; 1997. p 351-67.
50. Molenaar IW, Sijtsma K. User's manual MSP5 for Windows. Groningen: iecProGAMMA; 2000.
51. Sijtsma K, Molenaar IW. Introduction to nonparametric item response theory. Thousand Oaks: Sage; 2002.
52. Molenaar IW. Parametric and nonparametric item response theory models in health related quality of life measurements. In: Mesbah M, Cole BF, Lee ML, editors. *Statistical methods for quality of life studies: design, measurements and analysis*. Dordrecht: Kluwer Academic; 2002. p 143-54.
53. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.
54. Van Buuren S, Hopman-Rock M. Revision of the ICIDH Severity of Disabilities Scale by data linking and item response theory. *Stat Med* 2001;20:1061-76.
55. Roznowski M, Reith J. Examining the measurement quality of tests containing differentially functioning items: do biased items result in poor measurement? *Educ Psychol Meas* 1999;29:248-69.
56. Streiner DL, Norman GR. *Health Measurement Scales: a practical guide to their development and use*. 3rd ed. Oxford: Univ Pr; 2003.